AISR NAG5-10750: Spatial Statistics of Large Astronomical Databases: an Algorithmic Approach
Principal Investigator: István Szapudi

Final Report

In this AISRP, the we have demonstrated that the correlation function i) can be calculated for MAP in minutes (about 45 minutes for Planck) on a modest 500Mhz workstation ii) the corresponding method, although theoretically suboptimal, produces nearly optimal results for realistic noise and cut sky. This trillion fold improvement in speed over the standard maximum likelihood technique opens up tremendous new possibilities, which will be persued in the follow up.

1 Primary Results

During the first year we have tried several algorithms to calculate fast correlation function from CMB data. The runner up is an algorithm based on tree code. This has potential applications for the future, however, next we concentrate on our FFT based algorithm, which is the fastest to date.

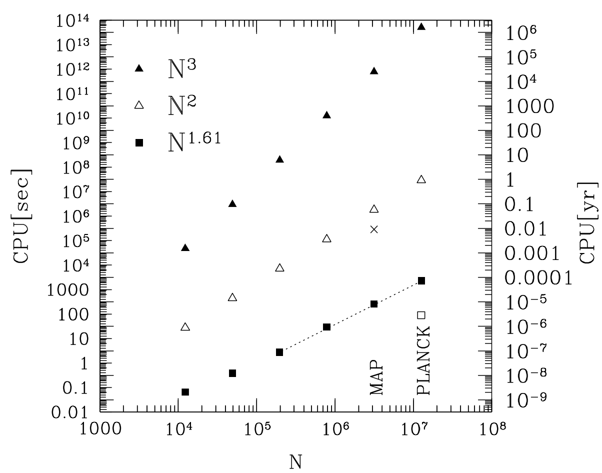

Figure 1: CPU requirements of selected CMB analysis methods are plotted: MADCAP (N^3, filled triangles, Borrill 1999), SPPSB (N^2, open triangles), and the present method ($\simeq N^{1.61}$, filled squares). The experiment specific technique of OSH for MAP is plotted with a cross, while an open square illustrates how Moore's law will shift the CPU for our method by 2007. All the squares are actual measured values on a 500 MHz CPU, while the triangles are based on extrapolation of scaling. The two leftmost points are relevant for MAP and Planck. The righthand y-axis displays CPU time in years, instead of seconds for clarity.

The recipe to extract C_ℓ's from a large CMB map is (SPPBS; Szapudi *et al.* 2001b): extract the two-point correlation function with an unbiased weighted estimator sampled at the roots of Legendre polynomials, then integrate with a Gauss-Legendre quadrature to obtain the C_ℓ's. Next we review the correlation function estimator.

Let us denote the temperature fluctuations at a sky vector n, a unit vector pointing to a pixel on the sky, with $T(n)$. In isotropic universes the two-point correlation function is a function only of the angle between the two vectors and can expanded into a Legendre series,

$$\xi_{12} = \langle T(n_1)T(n_2)\rangle = \sum_{\ell} \frac{\ell+1/2}{\ell(\ell+1)} \mathcal{C}_\ell P_\ell(\cos\theta), \tag{1}$$

where $\cos\theta = n_1.n_2$ is the dot product of the two unit vectors, $P_\ell(x)$ is the ℓ-th Legendre polynomial, and the $\mathcal{C}_\ell$ coefficients realize the angular power spectrum of fluctuations.

In reality each pixel value, Δ_i, contains contributions from the CMB and noise; the latter is also assumed to be Gaussian with a correlation matrix (the noise matrix) N_{ij}, determined during map-making. The full pixel-pixel correlation matrix is $C_{ij} = \xi_{ij} + N_{ij}$. matrix, produced as a sum of noise correlations measured in a set of noise realizations. Therefore we use the the estimator by Szapudi *et al.* (2001b)

$$\tilde{\xi}(\cos\theta) = \sum_{ij} f_{ij}(\Delta_i\Delta_j - \frac{1}{M}\sum_{k=1}^{M} n_i^k n_j^k), \tag{2}$$

where n_i^k is one of M realizations of the noise for pixel i. The key factors in gaining speed with the above estimator are heuristic weighting schemes, and efficient algorithms.

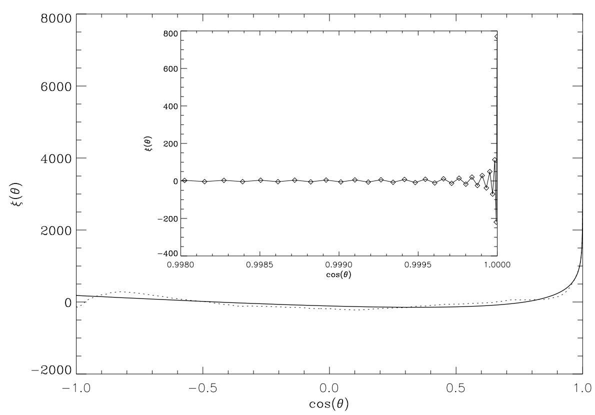

Figure 2: The figure displays the average correlation function in a simulated MAP survey (solid line) and one realization (dots). The inserted panel shows the average noise correlation function (line) with one realization (diamonds).

2 Summary

We have performed the basic algorithmic research underlying the goals of our project. This resulted in new correlation function engine, SpICE, (Spatially Inhomogenous Correlation Estimator), which can produce correlation functions and the angular power spectrum of the CMB in much shorter times then previously was possible. For the first time, accurate analysis of megapixel CMB maps is possible under very general assumptions on a simple workstation (or even notebook).

A version of the code is available at `http://www.ifa.hawaii.edu/users/szapudi/istvan.html` (click on AISR) in the format of a gzip compressed tar file. The accompanying `README` file gives enough information that the educated user will be able to compile and use the code. More improvements are planned for the future both in terms of functionality and user interface; the present code is capable of calculating the basic C_l's from HEALPix data.

The trillion fold improvement in speed over the standard maximum likelihood technique opens up tremendous new possibilities. In the next cycle we will continue with improvements and generalizations which will render the technique even more powerful: i) heuristic noise and signal weighting for the highest possible accuracy ii) generalization for polarization iii) generalization for higher order correlation functions to constrain the non-Gaussianity of the CMB as well as application to Sunyaev-Zeldovich effect and lensing iv) cross correlation of CMB with LSS and foregrounds v) component separation.

The relevant papers published in this cycle are listed below.

[1] Szapudi, I., Prunet, S., & Colombi, S. 2001, ApJLett, 561, 11

[2] Coil, Davis, & Szapudi 2001, PASP, 113, 131

[3] Szapudi, I., etal, (2002), ApJ, accepted